

Korrespondance analyse af markbilleder

Christoffer Thrysøe, Andreas Borgstad

Indholdsfortegnelse

1	Abstract	1
2	Introduction	2
2.1	Opgavens problemfelt	2
2.2	Problemformulering	2
3	Korrespondance Analyse	4
3.1	Detektor	4
3.2	Deskriptor	9
4	Analyse af markbilleder	9
4.1	Deskriptor	9
4.2	Krav til deskriptor	9
4.3	Matching	9
5	Analyse af markbilleder	10
5.1	Udfordringer	10
5.2	Krav til detektor	10
5.3	Krav til deskriptor	10
6	Gennemgang af detektionsmetoder	11
	Litteratur	12
A	Cost effectiveness Calculations	13

1 Abstract

To Be Done

2 Introduction

Future Cropping er et forsknings projekt, oprettet af Miljøstyrelsen, der i samarbejde med Data-logisk institut og institut for Plante- og Miljøvidenskab skal hjælpe landmænd i at automatisere detektion af ukrudt i marker. Projektet har til formål at nedsætte landbrugets brug af pesticider, ved hjælp af en drone ¹ udstyret med kamera og GPS. Dronen tager et antal overlappende luft billeder over landmandens mark, disse billeder bliver analyseret og omdannet til et ukrudtskort, tilknyttet GPS koordinater. Dette ukrudtskort behandles af en algoritme, der identificere præcis, hvor ukrudtet befinder sig og derved, hvor der er behov for pesticider[1].

2.1 Opgavens problemfelt

Opgavens afsæt i dette projekt er bestemmelsen af, hvordan de individuelle billeder passer sammen, hvilket er det første skridt i etableringen af et ukrudtskort. Dette udføres ved at etablere korrespondancer imellem billederne taget af dronen, hvilket er muligt da billederne overlapper hinanden. Denne teknik vil refereres til som "korrespondance analyse". Korrespondance analyse er en billedebehandlings teknik, indenfor "computer vision" og består af følgende trin. (1) Distinktive punkter detekteres i billederne som resultat af at anvende af en række matematiske modeller på billederne. Formålet ved dette stadie er udvælge de samme punkter/objekter i de overlappende billeder. (2) Området omkring disse fundne punkter beskrives af en deskriptor. (3) Hver deskriptor sammenlignes med andre fundne interessepunkter i andre billeder for at "matche" punkterne.

2.2 Problemformulering

Med udgangspunkt i litteraturen inden for korrespondanceanalyse samt implementering af flere eksisterende metoder, hvilke metoder, teoretisk og praktisk, anvendes bedst til korrespondanceanalyse af markbilleder?

Udvidelse af problemformuleringen

Der opstilles en beskrivelse af et udsnit af forskellige udvalgte metoder, der bliver brugt i feature detektion, feature deskription og matching. Der vil foretages korrespondanceanalyse af markbilleder, ved implementering af udvalgte eksisterende metoder.

Det eksperimentelle fokus i opgaven vil ligge på afprøvning af de forskellige metoder på markbilleder. Markbillederne har få distinktive træk, hvilket er egenskaber, der er vigtigt for korrespondance analyse. Udvælgelsen af metoder er baseret på hypoteser ift. hvordan metoderne forventes at reagere på markbilleder. <not done, empirisk>

Markbilledernes manglende diversitet, samt inkonsistente forhold under fotograferingen diskuteres i [ref til afsnit om markbillederne], præsenterer nogle potentielle udfordringer ved korrespondanceanalysen. Udfordringerne gør det derfor interessant at undersøge, hvorvidt det er muligt at finde tilpas nok punkter, der kan anvendes. Hovedproblemet ved opgaven fokuserer derfor på, hvordan der bedst kan detekteres features i markbillederne.

2.2.1 Afgrænsning

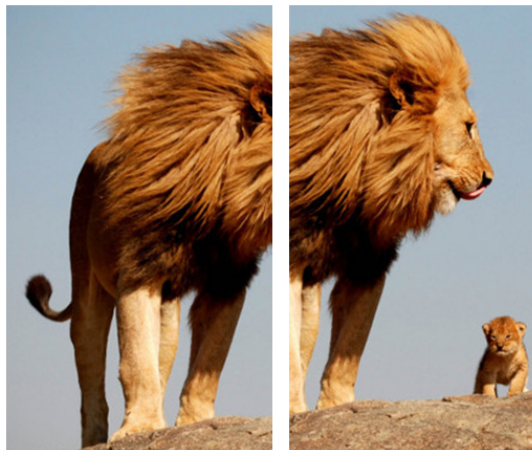
Projektet sigter på at afprøve allerede eksisterende metoder i et specifikt domæne og ikke skabe nye metoder. Programmet konstrueres mht. afklaring af de nævnte problemstillinger og ikke mhp. efterfølgende at blive anvendt i praksis. De udvalgte metoder implementeres mhp.

¹UAV (eng. unmanned aerial vehicle)

funktionaliteten, visse implementerings detaljer vil derfor undlades, hvis det enelige formål er at simplificere kompleksiteten.

3 Korrespondance Analyse

Korrespondance problemet, imellem to eller flere billeder af samme objekt, referere til problemet om at finde et sæt af punkter i det ene billede, der kan identificeres og matches i det andet. Udfordringen ved problemet ligger i at billederne, der skal matches, er udsat for en række ændringer, det kan f.eks. være forskydning i kameraets position ift. scenen eller ændringer i scenens motiv. Et godt eksempel på korrespondance problemet er det menneskelige syn. Øjnene agere som to kameraer, der hver især fanger deres billede og omdanner disse billeder til et sammenhængende panoramisk billede, ved hjælp af oprettelse af korrespondancer. Korrespondancen mellem øjnene tillader også opfattelse dybde i billedet, hvilket skyldes den horisontale forskydning af menneskets øjne. Denne sektion vil beskrive, hvordan denne korrespondance kan efterlignes af en computer



Figur 1: To billeder er af samme 2-D motiv, med kamerapositionen forskudt i x-aksen.

I figur 1 ses to billeder af det samme motiv, men hvor kameravinklen er forskudt. For at opnå en korrespondance imellem billederne skal der detekteres nogle unikke punkter, som skal optræde i begge billeder, dvs. der hvor billederne overlapper. Punkterne skal nøje beskrives så de kan genkendes i transformerede billeder. Til sidst skal punkterne matches for at estimere hvordan de korrespondere med hinanden, altså i dette tilfælde, hvordan de er forskudt i forhold til hinanden. Korrespondance analysens pipeline, der består af feature detektion, deskription og matching, er yderligere beskrevet herunder.

3.1 Detektor

Feature detektion er en metode indenfor billedbehandling, der determinere, for hvert punkt i et billede, om dette punkt er et interessepunkt. Interessepunktet er ofte fundet ved at evaluere et pixel område omkring det fundne punkt. Detektorens resultat vil derved være et subset af isolerede punkter fra billedet, der er markeret som interessante. (Så hvornår er punkter interessante?) Om et punkt er interessant, defineres udefra hvilke egenskaber der er interessante for applikationsdomænet. Det kan f.eks. være genkendelse af veje fra satellitbilleder, ved at finde kanter eller linjer. Interesse punkter defineres ikke udefra semantiske meningsfulde områder som ansigter eller objekter, da dette vil kræve en høj-niveau fortolkning af scenen. I stedet udvælges lokale pixel områder, der er matematisk distinktive, baseret på intensiteten i billedet. Der er mange måder at definere hvad et interessepunkt er og det er i sidste ende applikationsdomænet der afgør hvilke punkter der lokaliseres bedst, det kan dog overordnet defineres at en detektor skal besidde følgende egenskaber:

- *Repetierbar*: Givet to billeder, taget af samme objekt under forskellige betingelser, skal en stor del af de fundne punkter uafhængigt af hinanden, indgå i begge billeder. Lokaliseres

punkterne ikke tilstrækkeligt i begge billeder er det ikke muligt at oprette en korrespondance

- *Distinkte*: De fundne interessepunkter skal være unikke ift. intensitetsvariationen i det omkringliggende område. Dette muliggør bedere beskrivelse af punktet og derved større mulighed for korrekt match.

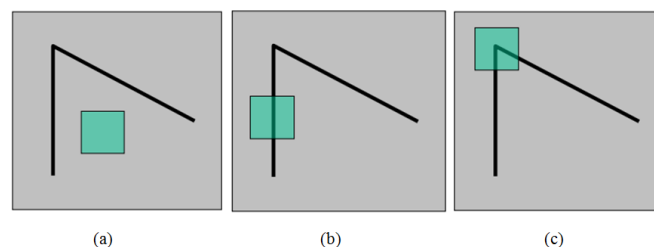
Disse egenskaber kan opnås ved følgende:

- *Invariants*: Detektionsmetoden skal tage højde for de ændringer der kan forekomme i mellem billederne, og derved være invariante overfor disse. Det kan f.eks være rotation af kameraet under billedtagning osv.
- *Robusthed*: Detektoren skal være robust overfor små deformationer som støj i billedet, hvilket fejlagtigt kan fortolkes som interessepunkter.

Udvælges punkterne ikke konsistent i begge billeder, er korrespondance oprettelsen ikke mulig. Nedenstående er en gennemgang af forskellige lokale mønstre som kan bruges til udvælgelse af interessepunkter.

3.1.1 Hjørner

En hjørnedetektor, leder efter hjørner i billedet og anvender disse som interessepunkter. Et hjørne kan defineres ved et punkt der har to dominerende kanter i hver sin retning.



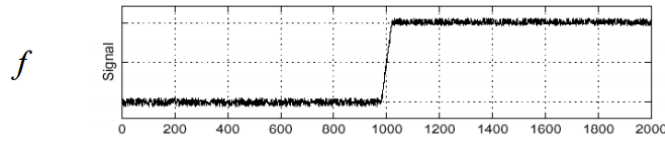
Figur 2: Tre udvalgte vinduer, med interessepunkter i centrum af samme motiv. (a) Punktet er lokaliseret i en teksturløs region, d.v.s. ingen teksturskift. (b) Punktet er lokaliseret på en kant. (c) Punktet er lokaliseret på et hjørne

Hvorfor er hjørner gode punkter? Et hjørne er et godt interessepunkt da der foregår store intensitetsskift omkring hjørnets omliggende område og er derfor distinkt ift. området. En intuitiv måde at definere hvorfor et punkt er interessant, er at placere et firkantet vindue omkring punktet. Dette vindue forskydes lokalt i x og y retningen. Resultere forskydningen af interessepunktet i et nyt objekt identisk med interessepunktet er punktet ikke lokalt distinkt. På figur 2 ses tre udvalgte punkter med et firkantet vindue placeret over. I stil med ovenstående definition, forskydes det firkantede vindue i alle retninger. Forskydes (a) vil det matche alle de forskudte billeder da regionen omkring punktet også er teksturløs. Punktet er derfor ikke distinkt. Forskydes (b) i x-aksen opnås et nyt objekt, men en forskydning i y-aksen vil resultere i samme objekt, og (b) er derfor ikke distinkt. Punktet placeret på et hjørne er distinkt da ingen forskydninger vil matche original billedet. Hjørnet kan derfor bruges som et distinktivt interessepunkt. At detektere hjørner er en udbredt teknik, da de er lokalt definerbare og ofte forekommer i forskellige scener. En matematisk definition kan opstilles og hele billedet og gennemses ift. definitionen.

3.1.2 Kanter

< intro gradient <http://www.slideshare.net/grim42/fuzzy-logic-based-edge-detection> > Kant detektion, referere til at finde kanter i billedet, hvor en kant kan defineres som et skarpt skift i billedintensiteten. Som nævnt er kanter ikke lokalt distinkte, men kan bruges til at fjerne en del

unødvendig information fra et billedet, ved kun at udtrykke kanterne. En måde at visualisere definitionen af en kant, er ved at opfatte billedet som et signal, en intensitetsfunktion, der afbilder billedintensiteten i 1-dimension. En høj kurve, angiver et skarpt intensitetsskift og derved en kant. Kanter kan derved identificeres ved at finde disse skarpe sving i intensitetsfunktionen.



Figur 3: En 1-dimensional fortolkning af intensiteten i et billede. De små udsving indikere støj, den store kurve repræsenter et skarpt skift i intensiteten og derved en kant i et billedet.

I figur 3 er en kant visualiseret efter ovenstående definition. Kanten i figuren kan findes ved at differentiere funktionen. En differentiering af funktionen vil angive, hvor skarp kurven er og derved fremhæve dens udsving. Et 2-dimensionelt billede er ikke en kontinuerlig funktion, og kan derved ikke differentieres i traditionel forstand, men består af diskrete værdier i form af pixel informationer. Differentiering af en funktion $f(x)$ kan approksimeres til følgende diskret differentiering:

$$\frac{df(x)}{dx} = \frac{f(x+1) - f(x-1)}{2} \quad (1)$$

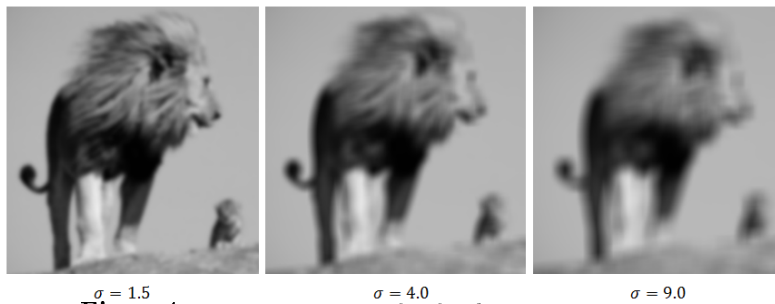
Ovenstående i (1) kan opnås ved at folde billedet med kernen $[1, 0, -1]$, hvor foldning af et billede I , med en kerne $K: I \oplus K$, hvor K har indgangen $m \times n$ og $I \oplus M \times N$ udregnes som:

$$O(i, j) = \sum_{k=1}^m \sum_{l=1}^m I(i+k-1, j+l-1) K(k, l) \quad (2)$$

Hvor det nye punkt i billedet er $O(i, j)$. Ved differentiering fremhæves udsving i intensitetsskiftene i billederne, hvilket gør det muligt at identificere kanterne. Problemet ved dette, visualiseret i figur 3, er at støj i billedet (de små udsving) også vil blive fremhævet, hvilket vil resultere i fejlagtige detektioner af kanter. For at imødegå dette foldes billedet med et gaussisk filter, hvilket er en approksimering til den gaussiske funktion. Den gaussiske funktion kan bruges, i billedbehandling, til at beskrive et punkt, ved en vægtet normalfordeling i de omliggende pixel, visualiseret som en klokke, der ligges over billedet. Foldningen med et gauss filter vil derfor resultere i en "flydende" overgang mellem pixels og derfor glatte billedet. Den gaussiske funktion i 2-D, hvor σ er standart afvigelsen, der determinere graden af glatning, er defineret i (3).

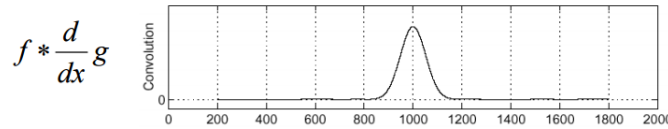
$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

Sigma værdien bestemmer, hvor meget et billede skal sløres. I figur 4 ses to billeder foldet med forskellige sigma værdier.



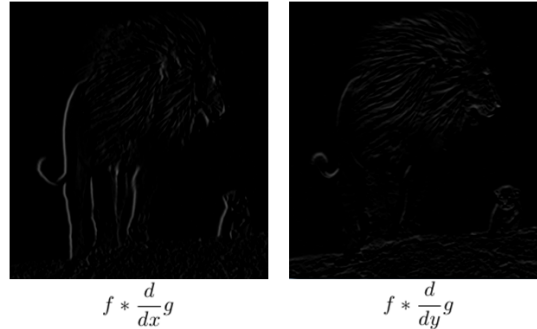
Figur 4: To billeder udsat for forskellige værdier i sigma.

For at undgå iteration af billedet to gange når kanter skal detekteres, for differentiering og sløring, kan dette udføres i en samlet operation ved at folde billedet med en differentieret gauss kerne, da foldning er en associativ operation,.



Figur 5: Resultatet af at folde et differentieret gauss filter på funktionen

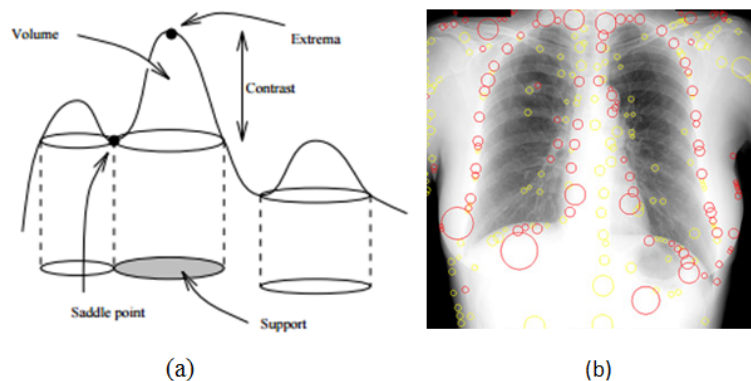
Kanterne er nu let definerbare som set i figur 6. I et 2-dimensionelt billede repræsenterer intensitetsskift også en orientering. Vertikale kanter findes ved at folde billedet med en gauss kerne differentieret i x-aksen og y-aksen for horisontale kanter som set i figur 6.



Figur 6: To billeder foldet med en differentieret gauss i x og y-aksen. Billedet viser tydeligt hvordan differentiering i x-aksen giver horisontale kanter og y-aksen giver vertikale kanter.

3.1.3 Blobs

En blob detektor detekterer regioner i billedet, hvor der sker en ændring af billedets egenskaber, det kan f.eks. være lysintensiteten ift. den omliggende region. En blob består af et sæt sammenhængende pixels alle af samme intensitet, der er distinktive ift. området. Lindenberg [2] definerer blobs som værende lyse regioner på sort baggrund eller omvendt, altså strukturer, der står i kontrast til deres baggrund. En blob kan derfor defineres som bestående af et område med *mindst* ét lokalt ekstremum, enten et maksimum eller et minimum. Blobbens struktur er ikke semantisk intuitivt definerbar, men er matematisk veldefineret <?>.



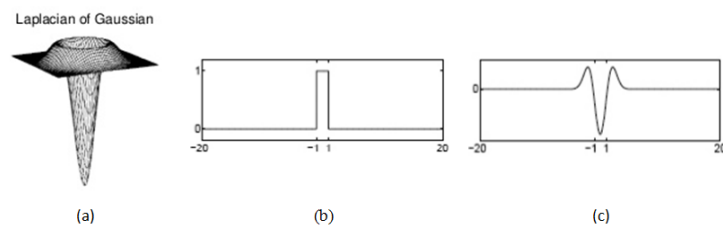
Figur 7: (a) En blob visualiseret i 2-d, udfra Lindenberg's definition. (b) En blob detektion udført på et røntgenbilledet af en brystkasse, cirklerne angiver fundne blobs, og størrelserne på disse definerer, skalaen af disse blobs.

I figur 7(a), ses en blob defineret af dets lokale ekstremum, hvor styrken af blobben beskrives ved kontrasten, ift. området omkring ekstremum. Lindenberg definerer blobbens domæne som værende afgrænset af dens "saddle point". Et "saddle point" angiver punktet, hvor intensiteten stopper med at falde og starter med at stige for lyse blobs, og modsat for mørke. Punktet definerer altså hvordan blobben er lokalt præsenteret. Denne definition på en blob stemmer overens med definitionen af interessepunkter. Det lokale ekstremum gør blobben vel defineret i regioner af billedet og derved distinktive. Blobs har også en fordel ift. kant og hjørne detektion, da blobs indgår i de fleste domæner undtagen homogene billeder. Kanter og hjørner forekommer ofte i menneskeskabte

objekter, ved veldefinerede strukturer. Blobs har derfor flere applikationsområder, og kan bruges indenfor billeder af objekter, der ikke er menneskelige defineret scener, f.eks. indenfor medicinsk billedanalyse til detektion af anomalier i røngten billeder som set i figur 7(b). Ligesom i kant-detektion, kan blobs beskrives ved intensitetsskift, hvor der forekommer en krusning rundt om ekstremaet. En metode til at detektere disse ekstremaer hedder Laplacian operatoren Δ^2 . Laplace operatoren anvendes på et gaussisk filter og opnår operatoren "Laplacian of Gaussian" (4):

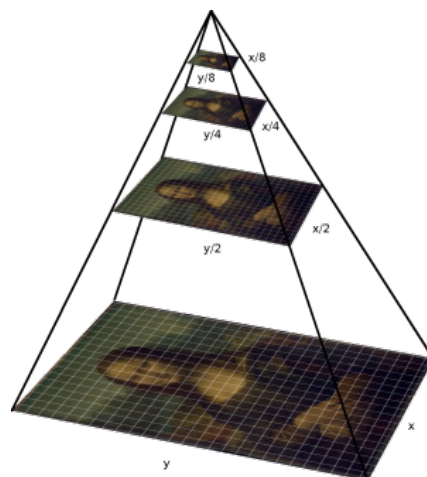
$$\Delta^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}, \Delta^2 G_\sigma(x, y) \quad (4)$$

Laplacian of gaussian operatoren kan diskritiseres og foldes med billedet for at finde lokale ekstremaer. Figur 8 viser, hvordan laplace operatoren opnår et maximum i centeret af en blob, og at blobben derved bliver lokaliserbar. Hvis blobben (b), er tyndere eller tykkere vil laplace operatoren ikke længere resultere i et ekstrema, men nærmere en kant. Det er derfor nødvendigt at normalisere laplace filteret ift. hvilken størrelse blobben har.



Figur 8: (a) En 3-D visualisering af en to-dimensional Laplacian of Gaussian (b) ét en-dimensionalt signal (c) Laplacian of Gaussian operatoren anvendt på (b)

Som illustreret i 7(b) forekommer blobs, forekommer blobs i forskellige størrelser. Ligesom i virkeligheden, har objekter forskellige strukturer ift. hvor langt væk objektet er og kan derfor optræde anderledes, hvis de opfattes på forskellige skalaer. Blob detection skal derfor foregå på forskelligt skalaniveau i billedet, for at opnå skala-invariants. Skala-rummet i et 2-dimensionalt billede repræsenteres af flere billeder i forskellige skalaer af det originale billede. Billeder, der repræsenterer forskellige skalaer, opnås ved at folde billedet med et 2-dimensionalt gaussisk filter som i (3).



Figur 9: En visualisering af et skala-rum formet som en pyramide af Leonardi Da Vinci's Mona Lisa. Hvert niveau angiver en skala repræsentation af det originale vindue, hvor toppen af pyramiden indeholder billeder af største skala og derfor med mindst information, og bunden af skalaen med det originale billede.

Et gaussisk filter bruges da gradvis højere værdier af σ fjerner strukturer, som vist i figur 4 og vigtigst af alt, at der ikke forekommer nye objekter ved transformationen fra finere til grovere skalaer [3]. Billedet, der repræsenterer flest detaljer, er den første skala. Idéen er derved at fjerne disse strukturer og fremhæve andre objekter gradvist på en større skala der også kan detekteres.

Detektionen af blobs sker altså på alle skalaer. Et billede i skala rummet for billedet $f(x, y)$ kan derfor defineres som i (5)

$$L(x, y, \sigma) = g(x, y, \sigma) * f(x, y) \quad (5)$$

hvor g er det 2-dimensionelle gaussiske filter, $L(x, y, \sigma)$ repræsenterer et billede i skal-rummet, og skala-parametren σ , bestemmer skalaen, eller placeringen i skal pyramiden. $L(x, y, 0) = f(x, y)$, da det er den "nederste" skala og den nederste del af skal pyramiden.

3.2 Deskriptor

4 Analyse af markbilleder

4.1 Deskriptor

Generelt er der mange måder, hvormed features kan repræsenteres: Et eksempel kunne være en hjørnedetektor, der som output giver en binær værdi, afhængigt af, om et punkt repræsenterer et hjørne eller ej - dette ville kunne repræsenteres, med et binært billede, hvor punkterne er enten 0 eller 1, afhængigt af, om de er et hjørne eller ej.

En robust repræsentation af en features kan ikke altid ske ved binære værdier. Der kan være brug for mere information: At tilføje en orientering af hjørnepunktet, samt intensiteten i punktet, ville øge informationen om hvert enkelt hjørne i billedet, hvilke kan være brugbart, relativt til anvendelsesområdet. Men generelt kommer forøgelsen af kompleksiteten af repræsentationen af en feature pris - der er mere data der skal bearbejdes.

Når der i denne opgave refereres til en feature deskriptor, menes her er en repræsentation af en feature - mere specifikt menes: En algoritme, der tager keypoints og et billede som input, og som output giver en repræsentation af featuren.

Af ofte brugte deskriptorer, kan her nævnes SIFT(fodnote) og SURF(fodnote). Førstnævnte skaber for hver feature, en vektor med 128 indgange, og sidstnævnte en vektor med 64 indgange. SURF, og specielt U-SURF har en klar fordel i forhold til beregningstid, specielt ved brug af integralbilleder, hvor køretidskompleksiteten på SIFT er højere, men tilgængelig er SIFT mere robust(fodnote). En god deskriptor afhænger af anvendelsesområdet, og her skal der laves kompromis mellem køretid og robusthed. Dette bliver beskrevet senere.

4.2 Krav til deskriptor

4.3 Matching

5 Analyse af markbilleder

5.1 Udfordringer

5.2 Krav til detektor

<Okklusion> Denne opgave vil forholde sig til problemet, der omhandler bevægelse af kameraet relativt til scenen og ikke ændringer i scenen.

OKKLUSION AF KORN

5.3 Krav til deskriptor

6 Gennemgang af detektionsmetoder

En matematisk formulering af hjørner kan opnås ved at bygge videre på notationen om et kvadratisk vindue omliggende det undersøgte punkt. De interessante punkter findes når forskydningen af vinduet resultere i nye objekter. Sammenlignes to vinduer, hvor begge portrætterer det samme objekt, vil intensitetsforskellen mellem disse vinduer være 0. Forskydningsvinduet kan derved bruges til at detektere hjørner, ved at tage et pixel område omkring et punkt, forskyde dette vindue i alle retninger og udregne forskellen. Dette kan beskrives med en vægtet funktion over vinduet (*weighted summed square difference*), der udregner forskellen på vinduerne:

$$E(u, v) = \sum_i (w_i, y_i) [I_0(x, y) - I_1(x + u, y + v)]^2 \quad (6)$$

hvor (u, v) er forskydningsvektoren, der bevæger sig mellem intervallet $[-1, 1]$. I_0 & I_1 er henholdsvis det originale vindue og det forskudte vindue. Forskellen, der resultere i den mindste forskel definere om punktet er interessant. Den mindste forskel tages da et punkt der f.eks er lokaliseret langs en kant resultere i et stort intensitetsskifte, når kanten krydses men ingen skift langs kanten. <moravec finder kanter hvis de er andet end horizontale eller vertikale wiki>

<moravec> <harris> <correclation> <laplace> <DOG> <DOH>

Litteratur

- [1] Droner kan mindske brug af pesticider, 2013. Københavns Universitet
[http : //www.science.ku.dk/presse/nyhedsarkiv/2013/droner _diku/](http://www.science.ku.dk/presse/nyhedsarkiv/2013/droner_diku/).
- [2] T. Lindeberg. *Detecting Salient Blob-Like Image Structures and Their Scales with a Scale-Space Primal Sketch - A Method for Focus-of-Attention*. Computational Vision and Active Perception Laboratory (CVAP), Department of Numerical Analysis and Computing Science, KTH (Royal Institute of Technology), 1993.
- [3] T. Lindeberg. *Feature Detection with Automatic Scale Selection*. Computational Vision and Active Perception Laboratory (CVAP), Department of Numerical Analysis and Computing Science, KTH (Royal Institute of Technology), 1998.

A Cost effectiveness Calculations

appendix test